



# Robust estimation in a nonlinear cointegration model

Jia Chen<sup>a,b</sup>, Degui Li<sup>a,b,\*</sup>, Lixin Zhang<sup>a</sup>

<sup>a</sup> Department of Mathematics, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> School of Economics, The University of Adelaide, Adelaide SA 5005, Australia

## ARTICLE INFO

### Article history:

Received 25 April 2008

Available online 17 September 2009

### AMS 2000 subject classifications:

62G05

62G20

62G35

### Keywords:

Cointegration model

Local time density

Nonparametric  $M$ -estimator

## ABSTRACT

This paper considers the nonparametric  $M$ -estimator in a nonlinear cointegration type model. The local time density argument, which was developed by Phillips and Park (1998) [6] and Wang and Phillips (2009) [9], is applied to establish the asymptotic theory for the nonparametric  $M$ -estimator. The weak consistency and the asymptotic distribution of the proposed estimator are established under mild conditions. Meanwhile, the asymptotic distribution of the local least squares estimator and the local least absolute distance estimator can be obtained as applications of our main results. Furthermore, an iterated procedure for obtaining the nonparametric  $M$ -estimator and a cross-validation bandwidth selection method are discussed, and some numerical examples are provided to show that the proposed methods perform well in the finite sample case.

© 2009 Elsevier Inc. All rights reserved.

## 1. Introduction

During the past two decades, there has been much interest in various nonparametric techniques to model time series data with possible nonlinearity. Both estimation and specification testing problems have been systematically examined for the case where the observed time series satisfy a type of stationarity. For recent development of them, we refer to [1–5] and the references therein.

As pointed out in the literature, the stationarity assumption seems too restrictive in practice. When tackling economic and financial issues from a time perspective, we often deal with nonstationary components. In reality, neither prices nor exchange rates follow a stationary distribution. Thus practitioners might feel more comfortable avoiding the stationary restriction. Until now, there has been a large amount of literature on nonparametric estimation of nonlinear regression and autoregression time series models with nonstationarity. The paper by Phillips and Park [6] was among the first to study nonparametric autoregression in the context of a random walk. Karlsen and Tjøstheim [7] and Karlsen et al. [8] independently discussed the nonparametric estimators in the framework of null recurrent Markov chains. Wang and Phillips [9] developed the asymptotic theory for local time density estimation and nonparametric cointegration models. For other recent development of nonparametric and semiparametric inferences in nonstationary time series, see [10–17] and the references therein.

The notion of cointegration was introduced by Granger [18] and Engle and Granger [19]. Two time series  $\{x_t\}$  and  $\{y_t\}$  are said to be linearly cointegrated if they are both nonstationary and if there exists a linear combination  $ax_t + by_t = w_t$  such that  $\{w_t\}$  is stationary. As the relationship between two time series is not necessarily linear, in this paper, we consider a nonlinear cointegration model defined by

$$y_t = m(x_t) + w_t, \quad (1.1)$$

\* Corresponding author at: Department of Mathematics, Zhejiang University, Hangzhou 310027, China.

E-mail addresses: [degui.li@adelaide.edu.au](mailto:degui.li@adelaide.edu.au), [ldgofzju@yahoo.com](mailto:ldgofzju@yahoo.com) (D. Li).

where  $m(\cdot)$  is some nonlinear function,  $\{x_t\}$  is some unit root type nonstationary input process defined as

$$x_t = x_{t-1} + v_t, \quad t \geq 1, \quad (1.2)$$

$x_0 = O_p(1)$ , and  $\{w_t\}$  and  $\{v_t\}$  are two sequences of stationary random variables satisfying some mild conditions.

A natural nonparametric estimator of  $m(z_0)$  in model (1.1) is the Nadaraya–Watson (NW) type estimator (cf. [8,9]),

$$\tilde{m}_n(z_0) = \frac{\sum_{t=1}^n y_t K\left(\frac{x_t - z_0}{h_n}\right)}{\sum_{t=1}^n K\left(\frac{x_t - z_0}{h_n}\right)}, \quad (1.3)$$

where  $K(\cdot)$  is some kernel function and  $h_n$  is the bandwidth. The NW estimator is widely used in nonparametric regression since its introduction. However, it is not robust due to the fact that the NW estimator can be considered as a local least-squares estimator and the least-squares estimator is not robust. For instance, it is sensitive to outliers and does not perform well when the error distribution is heavy-tailed. However, outliers or aberrant observations are observed very often in economic time series and finance, as well as in many other applied fields. A treatment of outliers or heavy-tailed errors is an important step in highlighting features of a data set. So in order to attenuate the lack of robustness of NW estimator,  $M$ -type regression estimator is a natural candidate for achieving desirable robustness properties. In Section 2, we will construct a robust version of nonparametric regression estimator for  $m(z_0)$ .

There is extensive literature on asymptotic properties of the robust nonparametric regression estimator in stationary time series. See, among others, [20] for locally weighted scatter plot smoothing, [21,22] for robust kernel method, [23] for smoothing spline methods, and [24] for local polynomial techniques. For recent development of nonparametric robust estimation, we refer to [25–28]. Recently, Lin et al. [29] studied the robust nonparametric estimation in nonstationary time series and established the asymptotic results in the framework of null recurrent Markov chains as in [7,8].

In this paper, we apply the local time limit theory to establish the asymptotic results of the nonparametric  $M$ -estimator. The local time approach was introduced by Phillips and Park [6] in the context of nonparametric autoregression and was developed recently by Wang and Phillips [9] for nonparametric cointegrating regression. The local time argument makes the approach in this paper more closely related to conventional nonparametric approaches than the null recurrent Markov chain method. In Section 2, we combine the local time argument with the nonparametric  $M$ -type smoothing technique to estimate the regression function  $m(\cdot)$ . Under some mild conditions, we obtain the weak consistency as well as the asymptotic distribution of the proposed estimator. As applications of our main results, we establish the asymptotic properties of local least square estimators (LLSE) and local least absolute distance estimators (LLADE). Since the fully iterative procedure for the nonparametric  $M$ -estimator is time consuming, we apply a one-step iterative procedure to reduce the computational burden. Furthermore, we discuss the bandwidth selection based on the robust cross-validation method and give some numerical examples to show that the nonparametric  $M$ -estimator performs well in finite sample case.

The rest of the paper is organized as follows. In Section 2, we give the definition of the nonparametric  $M$ -estimator and some assumptions. In Section 3, we state the asymptotic results together with some remarks. In Section 4, we provide an iterated procedure for the  $M$ -estimator, the choice of bandwidth as well as some numerical examples. In Section 5, we conclude the paper and give some extensions. In Appendix, we provide the proofs of the main results.

## 2. The estimation method and assumptions

The nonparametric  $M$ -estimator  $\hat{m}_n(z_0)$  of  $m(z_0)$  in (1.1) is defined as

$$\hat{m}_n(z_0) = \arg \min_{\theta} \sum_{t=1}^n \rho(y_t - \theta) K\left(\frac{x_t - z_0}{h_n}\right), \quad (2.1)$$

or as the solution to the equation

$$\sum_{t=1}^n \psi(y_t - \theta) K\left(\frac{x_t - z_0}{h_n}\right) = 0, \quad (2.2)$$

where  $\rho(\cdot)$  is a convex loss function,  $\psi(\cdot)$  is any choice of the subderivatives (or subgradient) of  $\rho(\cdot)$  and  $h_n$  is the bandwidth satisfying  $h_n \rightarrow 0$ . As the convex function  $\rho(\cdot)$  may not necessarily be differentiable at all points, we let  $\psi(\cdot)$  be any choice of its subderivatives. A subderivative of  $\rho(\cdot)$  at a point  $u_0$  is a real number  $c$  such that  $\rho(u) - \rho(u_0) \geq c(u - u_0)$  for all  $u$ . The set of all subderivatives of  $\rho(\cdot)$  at  $u_0$  is a nonempty closed interval  $[a, b]$  with  $a$  and  $b$  being the one-sided limits

$$a = \lim_{u \rightarrow u_0^-} \frac{\rho(u) - \rho(u_0)}{u - u_0}, \quad b = \lim_{u \rightarrow u_0^+} \frac{\rho(u) - \rho(u_0)}{u - u_0}.$$

Many authors have considered the asymptotic theory for  $M$ -estimators under some assumptions on the loss function  $\rho(\cdot)$  and its derivative  $\psi(\cdot)$ , such as [30–35]. However, most of these papers discussed particular choices of  $\rho(\cdot)$  and  $\psi(\cdot)$  (such as  $\rho(z) = |z|$ ), or general  $\rho(\cdot)$  and  $\psi(\cdot)$  under some restrictive conditions which do not cover some important special cases. Inspired by the paper by Bai et al. [36], here we allow  $\rho(\cdot)$  and  $\psi(\cdot)$  to include many commonly used estimators such as LLSE and LLADE.

The following assumptions will be made to establish the asymptotic properties of the nonparametric  $M$ -estimator.

- A1. The kernel  $K(\cdot)$  is nonnegative and has compact support, say  $[-1, 1]$ .  
 A2.  $\rho(\cdot)$  is a convex function and  $\psi(\cdot)$  is any choice of the subderivatives of  $\rho(\cdot)$ . Let  $\mathcal{D}$  be the set of discontinuity points of  $\psi(\cdot)$ , then  $P(\mathcal{D}) = 0$ .  
 A3. There exists a function  $\lambda_1(\cdot)$ , such that as  $|u| \rightarrow 0$ ,

$$E[\psi(w_t + u)|x_t = x] = \lambda_1(x)u + o(|u|), \quad \text{for all } 1 \leq t \leq n,$$

where  $\lambda_1(\cdot)$  is continuous at  $z_0$  with  $\lambda_1(z_0) > 0$ .

- A4. (i) Given  $\mathcal{F}_n(v) = \sigma\{v_t, 1 \leq t \leq n\}$ ,  $\{w_t, 1 \leq t \leq n\}$  is a sequence of independent and identically distributed (i.i.d.) random variables. Furthermore,

$$E[\psi^2(w_t)|x_t = x] = \sigma^2(x), \quad \text{for all } 1 \leq t \leq n,$$

where  $\sigma^2(x) > 0$  and  $\sigma^2(x)$  is continuous at  $z_0$ ;

(ii) Uniformly for  $x$  in a neighborhood of  $z_0$ ,

$$\sup_n \max_{1 \leq t \leq n} E[(\psi(w_t + u) - \psi(w_t))^2 | x_t = x] \leq \lambda(|u|),$$

where  $\lambda(\cdot)$  is continuous at 0 with  $\lambda(0) = 0$ ;

(iii) Uniformly for  $x$  in a neighborhood of  $z_0$ ,

$$\sup_n \max_{1 \leq t \leq n} E[|\psi(w_t)|^r | x_t = x] < \infty \quad \text{for some } r > 2.$$

- A5. The regression function  $m(\cdot)$  satisfies

$$|m(x + u) - m(x)| \leq C|u|^\beta, \quad \text{as } |u| \rightarrow 0 \text{ for some } \beta > 0.$$

- A6. Define the array  $\{x_{i,n}, 1 \leq i \leq n, n \geq 1\}$  by  $x_{i,n} = x_i/d_n$ , where  $\{d_n\}$  a sequence of positive numbers satisfying

$$d_n \rightarrow \infty \quad \text{and} \quad d_n/n \rightarrow 0.$$

Then, there exist a triple array of positive constants  $d_{l,k,n}$  and an array of  $\sigma$ -fields  $\mathcal{F}_{k,n}$  such that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \max_{0 \leq k \leq n-1} \frac{1}{n} \sum_{l=k+1}^n d_{l,k,n}^{-1} &< \infty, \\ \frac{1}{n} \sum_{i=1}^n d_{i,0,n}^{-1} &\rightarrow \tau, \quad \tau > 0, \end{aligned} \quad (2.3)$$

$x_{k,n}$  is adapted to  $\mathcal{F}_{k,n}$ , and given  $\mathcal{F}_{k,n}, (x_{l,n} - x_{k,n})/d_{l,k,n}, 0 \leq k < l \leq n$ , has a density  $\phi_{l,k,n}(\cdot)$  which is uniformly bounded and tends to a limit function  $\phi(\cdot)$  as  $n \rightarrow \infty$  and  $l - k \rightarrow \infty$ . Furthermore,  $\phi(\cdot)$  is continuous in a neighborhood of 0.

- A7. The bandwidth  $h_n$  satisfies

$$h_n \rightarrow 0, \quad nh_n/d_n \rightarrow \infty, \quad nh_n^{1+2\beta}/d_n \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

where  $\beta$  is defined as in A5.

**Remark 2.1.** The above assumptions are relatively mild in these kind of problems and can be justified in details. For example, A1 and A5 are quite natural and correspond to those used for the stationary time series. A2–A4 are assumed by many authors (see [37] for example) to establish the asymptotic properties of nonparametric  $M$ -estimators and they cover some well-known cases such as LLSE, LLADE and mixed LLSE and LLADE. This conditional independence assumption in A4 (i) is a little restrictive and it is satisfied when  $\{w_t\}$  is a sequence of i.i.d. random variables and is independent of  $\mathcal{F}_n(v) = \sigma\{v_t, 1 \leq t \leq n\}$ . A6 and A7 are similar to Assumption 2.3 in [9]. Next, we will give two examples in which A6 and A7 are satisfied.

**Example 2.1.** Assume that  $\{v_t\}$  is a sequence of i.i.d. random variables and  $v_t \sim N(0, 1)$ . It is easy to check that A6 is satisfied if we choose  $\mathcal{F}_{k,n} = \sigma\{v_1, \dots, v_k\}$ ,  $d_n = \sqrt{n}$  and  $d_{l,k,n} = \sqrt{l - k}/\sqrt{n}$ ,  $0 \leq k < l \leq n$ . In this case, A7 becomes

$$h_n \rightarrow 0, \quad \sqrt{n}h_n \rightarrow \infty, \quad \sqrt{n}h_n^{1+2\beta} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (2.4)$$

**Example 2.2.** Let  $\{v_t\}$  be a sequence of nonlinear transforms of a linear process, i.e.

$$v_t = H(\bar{v}_t), \quad \bar{v}_t = \sum_{k=0}^{\infty} b_k \eta_{t-k},$$

where  $\{\eta_i\}$  is a sequence of i.i.d. random variables satisfying

$$E\eta_1 = 0, \quad E\eta_1^2 = 1 \quad \text{and} \quad \int |\vartheta(t)|dt < \infty,$$

where  $\vartheta(\cdot)$  is the characteristic function of  $\eta_1$ . The above assumption is similar to the condition of Corollary 2.2 in [9]. Here,  $\{\bar{v}_t\}$  can be an ARMA or fractional ARIMA process. Wu [38] established functional limit theorems for the partial sums of  $\{\bar{v}_t\}$ . When  $H(v) = v$ , Wang and Phillips [9] showed that A6 and A7 hold under some mild conditions. Assume that

$$b_k \sim k^{-\alpha} L(k), \quad 1/2 < \alpha < 1,$$

where  $L(\cdot)$  is a slowly varying function and Condition 1 in [38] is satisfied.

Let  $p_0$  be the power rank of  $H(\cdot)$ , which is similar to the definition in [38]. If  $p_0(2\alpha - 1) < 1$ , by Theorem 1 in [38], we can show that A6 is satisfied with

$$\mathcal{F}_{k,n} = \sigma(\eta_i, -\infty < i \leq k), \quad d_n = C_{p_0} n^{1-p_0(\alpha-1/2)} L^{p_0}(n),$$

$C_{p_0}$  is some positive constant, and

$$d_{l,k,n}^2 = E(x_l - x_k)^2 / E x_n^2 = O\left(\frac{(l-k)^{3-2p_0\alpha} L^{2p_0}(l-k)}{n^{3-2p_0\alpha} L^{2p_0}(n)}\right).$$

In this case, A7 becomes

$$h_n \rightarrow 0, \quad n^{p_0(\alpha-1/2)} L^{-p_0}(n) h_n \rightarrow \infty, \quad \text{and} \quad n^{p_0(\alpha-1/2)} L^{-p_0}(n) h_n^{1+2\beta} \rightarrow 0, \quad (2.5)$$

as  $n \rightarrow \infty$ .

If  $p_0(2\alpha - 1) > 1$  or  $p_0(2\alpha - 1) = 1$  with  $\sum_{n=1}^{\infty} \frac{|L^{p_0}(n)|}{n} < \infty$ , by Theorem 1 in [38], we can show that A6 is satisfied with

$$\mathcal{F}_{k,n} = \sigma(\eta_i, -\infty < i \leq k), \quad d_n = B_{p_0} \sqrt{n},$$

$B_{p_0}$  is some positive constant, and

$$d_{l,k,n}^2 = E(x_k - x_l)^2 / E x_n^2 = O((l-k)/n).$$

Then, A7 is satisfied when (2.4) holds.

### 3. Asymptotic results

As our asymptotic theory relies heavily on the local time, we first introduce the local time of a continuous semimartingale  $\{M(s), s \geq 0\}$  at a point  $x$  as follows

$$L_M(t, x) = \lim_{\varepsilon \rightarrow 0} \frac{1}{2\varepsilon} \int_0^t 1\{x - \varepsilon < M(s) < x + \varepsilon\} d[M]_s, \quad (3.1)$$

where  $t$  and  $s$  are the time parameters,  $x$  is the spatial parameter, and  $[M]_s$  is the quadratic variation process of  $M(s)$  defined by

$$[M]_s = \lim_{\|T\| \rightarrow 0} \sum_{k=1}^m (X_{t_k} - X_{t_{k-1}})^2,$$

where  $T$  ranges over partitions of the interval  $[0, s]$  and the norm of the partition  $T$  is the mesh. In (3.1), the local time is defined with respect to the measure  $d[M]_s$ . If  $M(s)$  is a standard Brownian motion, then  $d[M]_s = ds$ . As can be seen from (3.1),  $L_M(t, x)$  measures how much time the process  $M(s)$  has spent at  $x$  up to time  $t$ . For more details about the local time, see [39,40].

The local time  $L_M(t, x)$  satisfies the following important Occupation Time Formula:

$$\int_0^t H(M(s)) d[M]_s = \int_{-\infty}^{+\infty} H(x) L_M(t, x) dx, \quad \text{for all } t \in \mathbf{R},$$

where  $H(\cdot)$  is a nonnegative transformation on  $\mathbf{R}$ .

Next, we present the asymptotic properties of the robust estimator  $\hat{m}_n(z_0)$ , which is defined in (2.1). We first give the consistency of  $\hat{m}_n(z_0)$  and then establish its asymptotic distribution.

**Theorem 3.1.** Assume that A1–A7 hold. Then there exists a solution  $\hat{m}_n(z_0)$  for (2.1) such that

$$\hat{m}_n(z_0) - m(z_0) = o_p(1). \quad (3.2)$$

**Theorem 3.2.** Assume that the conditions of Theorem 3.1 are satisfied and on a suitable probability space, there exists a stochastic process  $V(t)$  having a continuous local time  $L_V(t, s)$  such that

$$\sup_{0 \leq t \leq 1} |x_{[nt],n} - V(t)| = o_p(1), \quad (3.3)$$

where  $x_{i,n} = x_i/d_n$ ,  $[nt]$  is the integer part of  $nt$  and  $d_n$  is defined as in A6. Then, we have

$$(nh_n/d_n)^{\frac{1}{2}} (\hat{m}_n(z_0) - m(z_0)) \xrightarrow{d} (\sigma_1^2 L_V(1, 0))^{1/2} \xi, \quad (3.4)$$

where  $\sigma_1^2 = \sigma^2(z_0)v_2/(\lambda_1^2(z_0)\phi^2(0)v_1^2)$ ,  $v_j = \int K^j(u)du$ ,  $j = 1, 2$ ,  $\xi \sim N(0, 1)$  and  $\xi$  is independent of  $V$ ,  $L_V(t, s)$  stands for the local time of the process  $V(t)$  at the spatial point  $s$  over the time interval  $[0, t]$ .

**Remark 3.1.** The asymptotic distribution in Theorem 3.2 is mixed normal and is different from that for stationary time series (see [21,25] for example). (3.4) is also different from the asymptotic distribution obtained by Lin et al. [29]. By using the null recurrent Markov chain method, Lin et al. [29] established the asymptotic distribution of the local linear  $M$ -estimators with random convergence rate. When  $\{v_t\}$  is a sequence of i.i.d. random variables with  $Ev_1 = 0$  and  $Ev_1^2 = 1$ , it is easy to check that  $d_n = \sqrt{n}$ . Hence, the convergence rate in (3.4) is  $O(n^{1/4}h_n^{1/2})$ , which is the same as that obtained by Karlsen et al. [8] with  $\beta = 1/2$  in their paper and it is slower than the well-known convergence rate  $O(\sqrt{nh_n})$  for stationary nonparametric regression estimator.

**Remark 3.2.** The condition, that given  $\mathcal{F}_n(v)$ ,  $\{w_t\}$  is an independent sequence, can be weakened. It can be seen from the proof of Theorem 3.2 that this independence restriction can be relaxed to the condition that given  $\mathcal{F}_n(v)$ ,  $\{\psi(w_t)\}$  is a sequence of martingale differences. We conjecture that the results also hold for other dependent  $\{w_t\}$  such as  $\alpha$ -mixing or linear processes. But to focus on essentials in our development of asymptotic theory, we simply assume that  $\{w_t\}$  are independent given  $\mathcal{F}_n(v)$ . Furthermore, we can also establish analogous results for generalized nonparametric  $M$ -estimators such as local linear  $M$ -estimators or local polynomial  $M$ -estimators [25,26]. Since the proofs are similar, we will not give the details here.

**Remark 3.3.** Letting  $\rho(x) = x^2$  and  $\psi(x) = 2x$ , then estimator defined by (2.1) or (2.2) corresponds to the NW estimator or LLSE. Hence, we can obtain the asymptotic properties of the NW estimator as corollaries of the above two theorems. When  $\rho(x) = |x|$  and  $\psi(x) = \text{sign}(x)$ , the estimator (2.1) corresponds to LLAD. Assume that  $\{v_t\}$  is independent of  $\{w_t\}$  and denote by  $G(x)$  the distribution function of  $w_1$ . If  $G(x)$  has a density function  $g(x)$  around 0 and  $g(0) > 0$ , then we can also establish the asymptotic distribution for LLAD as a corollary of Theorem 3.2, i.e.,

$$(nh_n/d_n)^{1/2}(\tilde{m}_n(z_0) - m(z_0)) \xrightarrow{d} (\sigma_2^2 L_V(1, 0))^{1/2} \xi, \quad (3.5)$$

where  $\tilde{m}_n(z_0)$  is LLAD of  $m(z_0)$  and  $\sigma_2^2 = v_2/(g^2(0)\phi^2(0)v_1^2)$ . LLAD has been discussed by many authors in the stationary case, see, for instance, papers by Basset and Koenker [33], Bai et al. [35]. (3.5) is a new result for LLAD in nonstationary time series. Furthermore, Theorem 3.2 also holds for Huber's  $\psi$ -function.

#### 4. Examples of implementation

In this section, we discuss some critical problems such as the iterative algorithm for obtaining the nonparametric  $M$ -estimator of  $m(\cdot)$  and the choice of a proper bandwidth. We also experiment with two numerical examples to illustrate the proposed method.

##### 4.1. The iterated procedure and cross-validation bandwidth selection

The nonparametric  $M$ -estimator defined by (2.2) can be obtained by an iterative procedure. Define

$$\hat{\theta}_t(z_0) = \hat{\theta}_{t-1}(z_0) - (W_n(\hat{\theta}_{t-1}(z_0)))^{-1} \Psi_n(\hat{\theta}_{t-1}(z_0)), \quad (4.1)$$

where

$$\Psi_n(\theta) = \frac{d_n}{nh_n} \sum_{k=1}^n \psi(Y_k - \theta) K\left(\frac{x_k - z_0}{h_n}\right), \quad W_n(\theta) = \frac{\partial \Psi_n(\theta)}{\partial \theta},$$

The initial value of  $\hat{\theta}_0(z_0)$  can be arbitrarily chosen and the above procedure is terminated at  $t_0$ -th iteration if  $|\hat{\theta}_{t_0}(z_0) - \hat{\theta}_{t_0-1}(z_0)| < 0.0001$ . Then, we let the nonparametric  $M$ -estimator of  $m(z_0)$  to be  $\hat{\theta}_{t_0}(z_0)$ . However, this fully iterative procedure is time-consuming when the sample size is large. To overcome this disadvantage, we apply a one-step iterative procedure in our simulation as in [25]. The one-step  $M$ -estimator is defined by

$$\hat{\theta}_{OS}(z_0) = \hat{\theta}_0(z_0) - (W_n(\hat{\theta}_0(z_0)))^{-1} \Psi_n(\hat{\theta}_0(z_0)), \quad (4.2)$$

where  $\hat{\theta}_0(z_0)$  is the initial value. When the initial value satisfies

$$\hat{\theta}_0(z_0) - m(z_0) = O_P\left(\sqrt{\frac{d_n}{nh_n}}\right),$$

the one-step  $M$ -estimator  $\hat{\theta}_{OS}(z_0)$  has the same asymptotic properties as the  $M$ -estimator (see [25] for details). Following the arguments in [41,25], we can use the NW estimator of  $m(z_0)$  as an initial value. The simulation results below show that the one-step method works well in practice.

Another difficult problem in simulation is the choice of a proper bandwidth involved in nonparametric  $M$ -estimator. In this paper, we employ a robust cross-validation method as in [29]. The cross-validation method is very useful in assessing the performance of estimators via estimating their prediction errors. The basic idea is to set one of the data points aside for validation of a model and use the remaining data to build the model. Define

$$CV(h_n) = \sum_{t=1}^n (y_t - \hat{m}_{h_n, -t}(x_t))^2, \quad (4.3)$$

where  $\hat{m}_{h_n, -t}(x_t)$  is the nonparametric  $M$ -estimator with bandwidth  $h_n$  and the  $t$ -th observation left out. The bandwidth is selected to minimize  $CV(h_n)$ .

#### 4.2. Numerical examples

Next, we give two numerical examples to show that the nonparametric  $M$ -estimator is more robust than the NW estimator for contaminated observations.

**Example 4.1.** Consider the linear cointegration model

$$y_t = m(x_t) + w_t, \quad m(x) = x, \quad t = 1, 2, \dots, n, \quad (4.4)$$

where  $\{x_t\}$  is generated by the unit root process defined by

$$x_t = x_{t-1} + v_t, \quad v_t = 0.5v_{t-1} + \eta_t, \quad t \geq 1, \quad x_0 = 0, \quad v_0 \sim N(0, 4/3),$$

$\{\eta_t\}$  is a sequence of i.i.d. random variables with  $\eta_t \stackrel{i.i.d.}{\sim} N(0, 1)$ , and  $\{\eta_t\}$  is independent of  $\{w_t\}$ . The error  $\{w_t\}$  is independently taken from one of the following three distributions: (i) normal distribution  $N(0, \sigma^2)$ ,  $\sigma = 0.2$ , (ii) symmetric contaminated normal distribution  $0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2)$ , and (iii) Cauchy distribution  $0.1C(0, 1)$ .

The data from each of the above distributions consist of 1000 replications of samples of sizes  $n = 400, 800$  and  $1600$ . The uniform kernel  $K(u) = \frac{1}{2}I(|u| \leq 1)$  and Huber's  $\psi$ -function  $\psi(z) = \max\{-c, \min\{c, z\}\}$  with  $c = 1.25\sigma$  are applied in the simulation. The one-step iterative algorithm and the robust cross-validation bandwidth selection method defined in Section 4.1 are used to produce the nonparametric  $M$ -estimator of  $m(\cdot)$ . The measure of the performance of the estimators are taken to be the mean squared error (MSE) evaluated at the sample points  $x_1, \dots, x_n$ ,

$$MSE = \frac{1}{n} \sum_{t=1}^n (m_n(x_t) - m(x_t))^2, \quad (4.5)$$

where  $m_n(\cdot)$  is the  $M$ -estimator or the NW estimator of  $m(\cdot)$  and  $m(\cdot)$  is the true regression function. The simulation results are listed in Table 1. The quantities in Table 1 are the mean MSE's based on 1000 replications and the standard deviation of the MSE's throughout 1000 replications.

From Table 1, we can find that, for Example 4.1, both the nonparametric  $M$ -estimator and NW estimator perform well when the error  $\{w_t\}$  has the normal or contaminated norm distribution, and the performance of the two estimators improves as the sample size increases. However, when the error is heavy-tailed (Cauchy distributed), the nonparametric  $M$ -estimator behaves much better than NW estimator. The simulation results show that the nonparametric  $M$ -estimator is more robust than NW estimator.

**Example 4.2.** Consider the nonlinear cointegration model

$$y_t = m(x_t) + w_t, \quad m(x) = x^2, \quad t = 1, 2, \dots, n \quad (4.6)$$

where  $\{x_t\}$  is defined as in Example 4.1. We still study the following two cases: (i)  $\{w_t\}$  has a normal distribution  $N(0, \sigma^2)$ , and (ii)  $\{w_t\}$  has a symmetric contaminated normal distribution  $0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2)$ , and (iii) Cauchy distribution  $0.1C(0, 1)$ .

The data from each of the above distributions still consist of 1000 replications of samples of sizes  $n = 400, 800$  and  $1600$ . The results are reported in Table 2.

From Table 2, we can find that, for Example 4.2, the performance of the nonparametric  $M$ -estimator is much better than that of the NW estimator no matter which of the three distributions the error  $\{w_t\}$  take. Meanwhile, the performance of the two estimators in this example is worse than that in Example 4.1. This may due to the fact that when the regressor  $\{x_t\}$  is nonstationary, the volatility of  $\{x_t^2\}$  is much higher than that of  $\{x_t\}$  in Example 4.1.

## 5. Conclusions

In this paper, we establish the weak consistency as well as the asymptotic distribution for the proposed nonparametric  $M$ -estimator in a nonlinear cointegration model. We employ an iterated algorithm and a cross-validation bandwidth selection

**Table 1**

Simulation results for Example 4.1.

	NW estimator	M-estimator
	(i) $N(0, \sigma^2)$	
$n = 400$	0.0157 (0.0042)	0.0153 (0.0035)
$n = 800$	0.0108 (0.0027)	0.0108 (0.0026)
$n = 1600$	0.0081 (0.0021)	0.0082 (0.0021)
	(ii) $0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2)$	
$n = 400$	0.0162 (0.0040)	0.0161 (0.0037)
$n = 800$	0.0112 (0.0026)	0.0111 (0.0027)
$n = 1600$	0.0086 (0.0020)	0.0084 (0.0020)
	(iii) $0.1C(0, 1)$	
$n = 400$	$1.1473 \times 10^3$ ( $2.3899 \times 10^4$ )	3.6341 (36.2120)
$n = 800$	$2.0824 \times 10^3$ ( $6.0893 \times 10^4$ )	35.5750 ( $1.0334 \times 10^3$ )
$n = 1600$	$3.6817 \times 10^3$ ( $8.9410 \times 10^4$ )	2.9092 (71.8292)

**Table 2**

Simulation results for Example 4.2.

	NW estimator	M-estimator
	(i) $N(0, \sigma^2)$	
$n = 400$	16.4380 (16.7393)	0.8630 (0.6477)
$n = 800$	24.1650 (26.2842)	1.2600 (0.8324)
$n = 1600$	35.0640 (38.5541)	1.8181 (1.1684)
	(ii) $0.1N(0, 5^2\sigma^2) + 0.9N(0, \sigma^2)$	
$n = 400$	16.1361 (17.6536)	0.8649 (0.7106)
$n = 800$	22.3141 (24.7908)	1.2387 (0.8222)
$n = 1600$	35.5463 (45.6358)	1.7704 (1.1441)
	(iii) $0.1C(0, 1)$	
$n = 400$	$1.6787 \times 10^4$ ( $5.1600 \times 10^5$ )	6.1777 (86.0605)
$n = 800$	447.8955 ( $6.1013 \times 10^3$ )	33.2471 (812.8686)
$n = 1600$	$3.6730 \times 10^3$ ( $1.0790 \times 10^5$ )	15.6986 (282.2888)

method in the simulated examples. The simulation results show that the nonparametric  $M$ -estimator works well for both linear and nonlinear cointegration models even when the error is heavy-tailed.

There are many issues left for future study. For example, as suggested by the referees, we may allow for contemporaneous correlation between  $\{x_t\}$  and  $\{w_t\}$  as in [16]. In this case, we might have to apply a different method to establish the asymptotic theory. We may use Lemmas 7.1 and 7.2 in [16] to deal with the variance term in the proofs of Lemma A.1 and Theorem 3.1 in Appendix and apply the argument in the proof of Theorem 3.1 in [16] to obtain the asymptotic distribution of the nonparametric  $M$ -estimator. On the other hand, the limiting distribution of the proposed estimator will be different if the initial condition  $x_0 = O_P(1)$  is replaced by  $x_0 = O_P(n^{1/2})$  and this case will be considered in our future research. Other extensions include studying Bahadur representation of the proposed  $M$ -estimator as in [29]. Meanwhile, testing problems based on such robust estimation procedures will also be left for our future research.



## Acknowledgments

The authors would like to thank the Associate Editor and two anonymous referees for their valuable suggestions and comments, which greatly improved the former version of the paper. This work was supported by National Natural Science Foundation of China (Grant No. 10771192). The second author was also partially supported by the National Natural Science Foundation of China (Grant No. 10671176).

## Appendix

Before giving the proofs of [Theorems 3.1](#) and [3.2](#), we establish the following two lemmas which correspond to [Theorems 2.1](#) and [2.3](#) in [\[36\]](#). However, the proofs are more difficult here because of the nonstationarity of  $\{x_t\}$ . Throughout the proofs, we assume that  $x_0 = 0$  for simplicity. The basic idea also holds for the general case of  $x_0 = O_p(1)$  and the assumption of  $x_0 = O_p(1)$  will not affect the asymptotic properties (cf. [\[42\]](#)).

**Lemma A.1.** Denote  $\theta_0 = m(z_0)$  and  $K_t = K\left(\frac{x_t - z_0}{h_n}\right)$ ,  $t = 1, \dots, n$ . Suppose that all the conditions of [Theorem 3.1](#) hold, then for any constant  $c > 0$ , we have

$$\sup_{(nh_n/d_n)^{1/2}|\theta - \theta_0| \leq c} \left| \sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] - \frac{nh_n}{2d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 \right| = o_p(1). \quad (\text{A.1})$$

**Proof.** As  $\rho(\cdot)$  is convex, we have

$$\begin{aligned} K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] &= K_t [\psi(y_t - \theta_0 + \eta(\theta_0 - \theta)) - \psi(y_t - \theta_0)] (\theta - \theta_0) \\ &= K_t \{[\psi(y_t - \theta_0 + \eta(\theta_0 - \theta)) - \psi(w_t)] - [\psi(y_t - \theta_0) - \psi(w_t)]\} (\theta - \theta_0), \end{aligned} \quad (\text{A.2})$$

where  $|\eta| \leq 1$ . Letting

$$z_{t,n} = K_t [\psi(y_t - \theta_0) - \psi(w_t)] = K_t [\psi(w_t + m(x_t) - \theta_0) - \psi(w_t)]$$

and

$$z'_{t,n} = K_t [\psi(y_t - \theta_0 + \eta(\theta_0 - \theta)) - \psi(w_t)] = K_t [\psi(w_t + m(x_t) - \theta_0 + \eta(\theta_0 - \theta)) - \psi(w_t)],$$

we have

$$K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] = (\theta - \theta_0)(z_{t,n} - z'_{t,n}). \quad (\text{A.3})$$

Hence, it is easy to check that

$$\begin{aligned} \text{Var} \left( \sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] \right) &= (\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^n (z_{t,n} - z'_{t,n}) \right) \\ &\leq 2(\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^n z_{t,n} \right) + 2(\theta - \theta_0)^2 \text{Var} \left( \sum_{t=1}^n z'_{t,n} \right). \end{aligned} \quad (\text{A.4})$$

Noting that for  $|x_t - z_0| \leq h_n$  and by [A5](#), we have

$$|m(x_t) - \theta_0| = |m(x_t) - m(z_0)| \leq Ch_n^\beta, \quad (\text{A.5})$$

$$|m(x_t) - \theta_0 + \eta(\theta_0 - \theta)| \leq |m(x_t) - \theta_0| + |\theta - \theta_0| \leq Ch_n^\beta + C(nh_n/d_n)^{-1/2}. \quad (\text{A.6})$$

By [\(A.5\)](#), [A4\(ii\)](#), [A6](#) and [A7](#), we know that for  $n$  sufficiently large,

$$\begin{aligned} \text{Var} \left( \sum_{t=1}^n z_{t,n} \right) &= \sum_{t=1}^n \text{Var}(z_{t,n}) \leq \sum_{t=1}^n E z_{t,n}^2 \\ &= \sum_{t=1}^n E \left( K_t^2 E[(\psi(w_t + m(x_t) - \theta_0) - \psi(w_t))^2 | x_t] \right) \\ &\leq \sum_{t=1}^n E \left( K_t^2 \lambda(m(x_t) - \theta_0) \right) \end{aligned}$$



$$\begin{aligned}
&= o\left(h_n d_n^{-1} \sum_{t=1}^n d_{t,0,n}^{-1} \int K^2(u) \phi_{t,0,n}\left(\frac{h_n u + z_0}{d_n d_{t,0,n}}\right) du\right) \\
&= o\left(h_n d_n^{-1} \sum_{t=1}^n d_{t,0,n}^{-1}\right) = o(nh_n/d_n).
\end{aligned} \tag{A.7}$$

By (A.6) and applying the same argument as (A.7), we can obtain

$$\text{Var}\left(\sum_{t=1}^n z'_{t,n}\right) = o(nh_n/d_n). \tag{A.8}$$

From (A.4), (A.7) and (A.8) we know that for  $(nh_n/d_n)^{1/2}|\theta - \theta_0| \leq c$

$$\text{Var}\left(\sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)]\right) = o(1). \tag{A.9}$$

On the other hand, by A3 and the convexity of  $\rho(\cdot)$ , we can show that

$$E[\rho(w_t + u) - \rho(w_t)|x_t = x] = \frac{1}{2}\lambda_1(x)u^2 + o(|u|^2). \tag{A.10}$$

The proof of (A.10) is similar to Lemma 1 in [36]. As a result,

$$\begin{aligned}
E\left(\sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0)]\right) &= E\left(\sum_{t=1}^n \frac{1}{2} K_t \lambda_1(x_t) [(m(x_t) - \theta)^2 - (m(x_t) - \theta_0)^2]\right) \\
&= \frac{1}{2}(\theta - \theta_0) \sum_{t=1}^n E[K_t \lambda_1(x_t)(\theta + \theta_0 - 2m(x_t))]
\end{aligned} \tag{A.11}$$

and

$$E\left(\sum_{t=1}^n K_t \psi(y_t - \theta_0)(\theta - \theta_0)\right) = (\theta - \theta_0) \sum_{t=1}^n E[K_t \lambda_1(x_t)(m(x_t) - \theta_0)]. \tag{A.12}$$

By (2.3), (A.11) and (A.12), we obtain

$$\begin{aligned}
E\left(\sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)]\right) &= \frac{1}{2}(\theta - \theta_0)^2 \sum_{t=1}^n E(K_t \lambda_1(x_t)) \\
&= \frac{1}{2}(\theta - \theta_0)^2 \lambda_1(z_0) h_n d_n^{-1} \sum_{t=1}^n d_{t,0,n}^{-1} \times \phi(0) \int K(u) du (1 + o(1)) \\
&= \frac{1}{2}(\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) v_1 n h_n d_n^{-1} (1 + o(1)).
\end{aligned} \tag{A.13}$$

By (A.9) and (A.13), we have

$$\sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] - \frac{1}{2}(\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) v_1 \frac{nh_n}{d_n} = o_p(1).$$

Since

$$\sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)]$$

is convex in  $\theta$ , and  $\frac{nh_n}{2d_n}(\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) v_1$  is convex and continuous in  $\theta$ , (A.1) is proved by Theorem 10.8 in [43].  $\square$

**Lemma A.2.** Under the conditions of Theorem 3.1, we have, for any  $c > 0$ ,

$$\sup_{(nh_n/d_n)^{1/2}|\theta - \theta_0| \leq c} \left| \left(\frac{nh_n}{d_n}\right)^{-\frac{1}{2}} \sum_{t=1}^n K_t [\psi(y_t - \theta) - \psi(y_t - \theta_0)] + \left(\frac{nh_n}{d_n}\right)^{\frac{1}{2}} (\theta - \theta_0) \lambda_1(z_0) v_1 \phi(0) \right| = o_p(1).$$

**Proof.** Using the same method as that in the proof of Lemma A.1 and by Theorem 2.5.7 in [43], we can prove Lemma A.2. Details are omitted here.  $\square$

**Proof of Theorem 3.1.** Let  $\widehat{\theta}_n = \widehat{m}_n(z_0)$ . We first prove

$$(nh_n/d_n)^{1/2}(\widehat{\theta}_n - \theta_0) = O_P(1).$$

It suffices for us to prove that, for any positive sequence  $\{c_n\}$  satisfying  $c_n \rightarrow \infty$ ,

$$P\left((nh_n/d_n)^{1/2}|\widehat{\theta}_n - \theta_0| \geq c_n\right) \rightarrow 0. \quad (\text{A.14})$$

By (A.1), we can choose a sequence of positive numbers  $\{c'_n\}$ , such that  $c'_n \rightarrow \infty$ ,  $c'_n \leq c_n$  and

$$\begin{aligned} & \sup_{(nh_n/d_n)^{1/2}|\theta - \theta_0| \leq c'_n} \left| \sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0) + \psi(y_t - \theta_0)(\theta - \theta_0)] - \frac{nh_n}{2d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 \right| \\ & = o_P(1). \end{aligned} \quad (\text{A.15})$$

When  $(nh_n/d_n)^{1/2}|\theta - \theta_0| = c'_n$ , we have

$$\frac{nh_n}{2d_n} (\theta - \theta_0)^2 \lambda_1(z_0) \tau \phi(0) \nu_1 \geq \frac{1}{4} \lambda_1(x_0) \tau \phi(0) \nu_1 c_n'^2. \quad (\text{A.16})$$

Next, we will prove

$$\sum_{t=1}^n K_t \psi(y_t - \theta_0) = O_P\left((nh_n/d_n)^{1/2}\right). \quad (\text{A.17})$$

Note that

$$\begin{aligned} E\left(\sum_{t=1}^n K_t \psi(y_t - \theta_0)\right)^2 &= \sum_{t=1}^n E(K_t \psi(y_t - \theta_0))^2 + 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n E(K_t K_j \psi(y_t - \theta_0) \psi(y_j - \theta_0)) \\ &=: J_{n1} + J_{n2}. \end{aligned} \quad (\text{A.18})$$

From (A.7) and A4(i), we have

$$\begin{aligned} J_{n1} &= \sum_{t=1}^n E(K_t[(\psi(y_t - \theta_0) - \psi(w_t)) + \psi(w_t)])^2 \\ &\leq 2 \sum_{t=1}^n E[K_t(\psi(y_t - \theta_0) - \psi(w_t))]^2 + 2 \sum_{t=1}^n E[K_t \psi(w_t)]^2 \\ &= 2 \sum_{t=1}^n E z_{t,n}^2 + 2 \sum_{t=1}^n E(K_t^2 \sigma^2(x_t)) \\ &= o(nh_n/d_n) + O\left(\sum_{t=1}^n E K_t^2\right) \\ &= o(nh_n/d_n) + O(nh_n/d_n) = O(nh_n/d_n). \end{aligned} \quad (\text{A.19})$$

By A4 (i), we have given  $x_i$  and  $x_j$ ,  $w_i$  and  $w_j$  are independent. Therefore, by A3, A6 and (A.5), we have

$$\begin{aligned} J_{n2} &= 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n E(K_t K_j E[\psi(w_t + m(x_t) - \theta_0) \psi(w_j + m(x_j) - \theta_0) | x_t, x_j]) \\ &= 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n E(K_t K_j E[\psi(w_t + m(x_t) - \theta_0) | x_t] E[\psi(w_j + m(x_j) - \theta_0) | x_j]) \\ &= 2 \sum_{t=1}^{n-1} \sum_{j=t+1}^n E(K_t K_j \lambda_1(x_t) \lambda_1(x_j) (m(x_t) - \theta_0) (m(x_j) - \theta_0)) \\ &= O\left(h_n^{2\beta} \sum_{t=1}^{n-1} \sum_{j=t+1}^n E K_t K_j\right) \end{aligned}$$

$$\begin{aligned}
&= O \left( h_n^{2\beta} \sum_{t=1}^{n-1} \sum_{j=t+1}^n E \left[ K_t \int K \left( \frac{d_{j,t,n} d_n u + x_t - z_0}{h_n} \right) \phi_{j,t,n}(u) du \right] \right) \\
&= O \left( h_n^{1+2\beta} d_n^{-1} \sum_{t=1}^{n-1} E K_t \sum_{j=t+1}^n d_{j,t,n}^{-1} \right) \\
&= O \left( n h_n^{1+2\beta} d_n^{-1} \sum_{t=1}^n E K_t \right) \\
&= O \left( n^2 h_n^{2+2\beta} d_n^{-2} \right) \\
&= o(nh_n/d_n).
\end{aligned} \tag{A.20}$$

The last equality holds since  $nh_n^{1+2\beta}/d_n \rightarrow 0$ . In view of (A.18)–(A.20), we obtain (A.17). Therefore, when  $(nh_n/d_n)^{1/2}|\theta - \theta_0| = c'_n$ ,

$$\left| (\theta - \theta_0) \sum_{t=1}^n K_t \psi(y_t - \theta_0) \right| = O_P(c'_n). \tag{A.21}$$

As  $c'_n \rightarrow \infty$ , we know, from (A.15), (A.16) and (A.21),

$$P \left( \inf_{(nh_n/d_n)^{1/2}|\theta - \theta_0| = c'_n} \sum_{t=1}^n K_t [\rho(y_t - \theta) - \rho(y_t - \theta_0)] \leq 0 \right) \rightarrow 0.$$

By the convexity of  $\rho(\cdot)$ , we have

$$P \left( \inf_{(nh_n/d_n)^{1/2}|\theta - \theta_0| \geq c'_n} \sum_{t=1}^n K_t \rho(y_t - \theta) \leq \sum_{t=1}^n K_t \rho(y_t - \theta_0) \right) \rightarrow 0.$$

By the definition of  $\hat{\theta}_n$  in (2.1), we get

$$P \left( (nh_n/d_n)^{1/2} |\hat{\theta}_n - \theta_0| \geq c'_n \right) \rightarrow 0,$$

which implies (A.14) since  $c'_n \leq c_n$ . So  $(nh_n/d_n)^{1/2}(\hat{\theta}_n - \theta_0) = O_P(1)$  holds. Reminding  $nh_n/d_n \rightarrow \infty$ , we know that (3.3) is valid.  $\square$

**Proof of Theorem 3.2.** We have proved, in the proof of Theorem 3.1, that

$$(nh_n/d_n)^{1/2}(\hat{\theta}_n - \theta_0) = O_P(1).$$

By Lemma A.2, we have

$$\left( \frac{nh_n}{d_n} \right)^{-1/2} \sum_{t=1}^n K_t [\psi(y_t - \hat{\theta}_n) - \psi(y_t - \theta_0)] + \left( \frac{nh_n}{d_n} \right)^{1/2} (\hat{\theta}_n - \theta_0) \lambda_1(z_0) v_1 \phi(0) = o_P(1).$$

Noting that  $\sum_{i=1}^n K_i \psi(y_i - \hat{\theta}_n) = 0$  and from the above equation, we have

$$\hat{\theta}_n - \theta_0 = (nh_n/d_n)^{-1} (\lambda_1(z_0) v_1 \phi(0))^{-1} \sum_{i=1}^n K_i \psi(y_i - \theta_0) + o_P((nh_n/d_n)^{-1/2}). \tag{A.22}$$

Note that

$$\sum_{i=1}^n K_i \psi(y_i - \theta_0) = \sum_{i=1}^n K_i \psi(w_i) + \sum_{i=1}^n z_{i,n}. \tag{A.23}$$

From (A.7), we know that  $\sum_{i=1}^n E z_{i,n}^2 = o(nh_n/d_n)$ . On the other hand, by taking the same lines as (A.20), we can get  $\sum_{i=1}^{n-1} \sum_{j=i+1}^n E z_{i,n} z_{j,n} = o(nh_n/d_n)$ . As a result,

$$E \left( \sum_{i=1}^n z_{i,n} \right)^2 = o(nh_n/d_n),$$

which implies

$$\sum_{i=1}^n z_{i,n} = o_p((nh_n/d_n)^{1/2}). \quad (\text{A.24})$$

By (A.22)–(A.24), we have

$$\hat{\theta}_n - \theta_0 = (nh_n/d_n)^{-1}(\lambda_1(z_0)v_1\phi(0))^{-1} \sum_{i=1}^n K_i\psi(w_i) + o_p((nh_n/d_n)^{-1/2}). \quad (\text{A.25})$$

As  $K(\cdot)$  is an integrable function, by Theorem 2.1 in [9] we get

$$(nh_n/d_n)^{-1} \sum_{i=1}^{[nr]} K_i^2 \xrightarrow{P} L_V(1, 0) \int K^2(u)du = L_V(1, 0)v_2. \quad (\text{A.26})$$

On the other hand, by A4(i), we know that, given  $\mathcal{F}_n(v)$ ,  $\{K_i\psi(w_i), 1 \leq i \leq n\}$  is a sequence of martingale differences. Therefore, by A4(i), A8, (A.26) and the central limit theorem for martingale differences [44], we know

$$(nh_n/d_n)^{-1/2} \sum_{i=1}^n K_i\psi(w_i) \xrightarrow{d} (\sigma^2(z_0)L_V(1, 0)v_2)^{1/2} \xi, \quad (\text{A.27})$$

where  $\xi \sim N(0, 1)$  and  $\xi$  is independent of  $V$ . In view of (A.25) and (A.27), the proof of Theorem 3.2 is completed.  $\square$

## References

- [1] P.J. Green, B.W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall, London, 1994.
- [2] J. Fan, I. Gijbels, *Local Polynomial Modelling and its Applications*, Chapman & Hall, London, 1996.
- [3] J. Fan, Q. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York, 2003.
- [4] J. Gao, *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman & Hall/CRC, London, 2007.
- [5] Q. Li, J. Racine, *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, Princeton, 2007.
- [6] P.C.B. Phillips, J.Y. Park, Nonstationary density estimation and kernel autoregression, Cowles Foundation Discussion Paper No. 1181 (1998).
- [7] H.A. Karlsen, D. Tjøstheim, Nonparametric estimation in null recurrent time series, *Ann. Statist.* 29 (2001) 372–416.
- [8] H.A. Karlsen, T. Mykkelbust, D. Tjøstheim, Nonparametric estimation in a nonlinear cointegration type model, *Ann. Statist.* 35 (2007) 252–299.
- [9] Q.Y. Wang, P.C.B. Phillips, Asymptotic theory for local time density estimation and nonparametric cointegrating regression, *Econom. Theory* 25 (2009) 710–738.
- [10] U.K. Müller, G. Elliott, Tests for unit roots and the initial condition, *Econometrics* 71 (2003) 1269–1286.
- [11] G. Elliott, U.K. Müller, Minimizing the impact of the initial condition on testing for unit roots, *J. Econometrics* 135 (2006) 285–310.
- [12] M. Schienle, Nonparametric nonstationary regression, Working Paper of Mannheim University (2008).
- [13] Z. Cai, Q. Li, J.Y. Park, Functional-coefficient models for nonstationary time series data, *J. Econometrics* 148 (2009) 101–113.
- [14] J. Chen, J. Gao, D. Li, Semiparametric regression in null recurrent time series, 2009, Working Paper available at [www.adelaide.edu.au/directory/jiti.gao](http://www.adelaide.edu.au/directory/jiti.gao).
- [15] J. Gao, M. King, Z. Lu, D. Tjøstheim, Specification testing in nonlinear and nonstationary time series regression, *Ann. Statist.* (2009) (forthcoming).
- [16] Q.Y. Wang, P.C.B. Phillips, Structural nonparametric cointegrating regression, *Econometrica* (2009) (forthcoming).
- [17] Q.Y. Wang, P.C.B. Phillips, Asymptotic theory for zero density estimation with nonparametric regression applications, Cowles Foundation Discussion Paper No. 1687 (2009).
- [18] C.W.J. Granger, Some properties of time series data and their use in econometric model specification, *J. Econometrics* 16 (1981) 121–130.
- [19] R. Engle, C.W.J. Granger, Cointegration and error correction: Representation, estimation and testing, *Econometrica* 55 (1987) 251–276.
- [20] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, *J. Amer. Statist. Assoc.* 74 (1979) 829–836.
- [21] W. Härdle, Robust regression function estimation, *J. Multi. Anal.* 14 (1984) 169–180.
- [22] G. Boente, R. Fraiman, Robust nonparametric regression estimation for dependent observations, *Ann. Statist.* 17 (1989) 1242–1256.
- [23] J.K. Cunningham, R.L. Eubank, T. Hsing,  $M$ -type smoothing splines with auxiliary scale estimation, *Comput. Statist. Data Anal.* 36 (1991) 115–125.
- [24] A.H. Welsh, Robust estimation of smooth regression and spread functions and their derivatives, *Statist. Sin.* 6 (1994) 347–366.
- [25] J. Fan, J. Jiang, Variable bandwidth and one-step local  $M$ -estimator, *Sci. China Ser. A* 43 (2000) 65–81.
- [26] J. Jiang, Y.P. Mack, Robust local polynomial regression for dependent data, *Statist. Sin.* 11 (2001) 705–722.
- [27] J. Beran, S. Ghosh, P. Sibbertsen, Nonparametric  $M$ -estimation with long-memory errors, *J. Statist. Plann. Inference* 117 (2003) 199–205.
- [28] Z. Cai, E. Ould-Saïd, Local  $M$ -estimator for nonparametric time series, *Statist. Probab. Lett.* 65 (2003) 433–449.
- [29] Z. Lin, D. Li, J. Chen, Local linear  $M$ -estimators in null recurrent time series, *Statist. Sin.* 19 (2009a) 1683–1703.
- [30] P.J. Huber, Robust estimation of a location parameter, *Ann. Math. Statist.* 35 (1964) 73–101.
- [31] P.J. Huber, Robust regression, *Ann. Statist.* 1 (1973) 799–821.
- [32] P.J. Bickel, One-step Huber estimates in the linear model, *J. Amer. Statist. Assoc.* 70 (1975) 428–433.
- [33] G. Basset, R. Koenker, Asymptotic theory of least absolute error regression, *J. Amer. Statist. Assoc.* 73 (1978) 618–622.
- [34] S. Heiler, R. Willers, Asymptotic normality of  $R$ -estimates in the linear model, *Statistics* 19 (1988) 173–184.
- [35] Z.D. Bai, X.R. Chen, Y. Wu, L.C. Zhao, Asymptotic normality of minimum  $L_1$ -norm estimates in linear models, *Sci. China Ser. A* 33 (1990) 449–463.
- [36] Z.D. Bai, C.R. Rao, Y. Wu,  $M$ -estimation of multivariate linear regression parameters under a convex discrepancy function, *Statist. Sin.* 2 (1992) 237–254.
- [37] Z. Lin, D. Li, J. Gao, Local linear  $M$ -estimation in nonparametric spatial regression, *J. Time Ser. Anal.* 30 (2009b) 286–314.
- [38] W.B. Wu, Unit root testing for functionals of linear processes, *Econom. Theory* 22 (2006) 1–14.
- [39] D. Geman, J. Horowitz, Occupation densities, *Ann. Probab.* 8 (1980) 1–67.
- [40] D. Revuz, M. Yor, *Continuous Martingale and Brownian Motion*, 2nd ed., Springer-Verlag, New York, 1994.
- [41] J. Fan, J. Chen, One-step local quasi-likelihood estimation, *J. Roy. Statist. Soc. B* 61 (1999) 927–943.
- [42] J.Y. Park, P.C.B. Phillips, Asymptotics for nonlinear transformations of integrated time series, *Econom. Theory* 15 (1999) 269–298.
- [43] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [44] P. Hall, C.C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, New York, 1980.